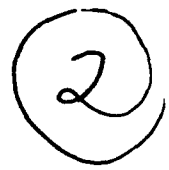
1



CORTICON INC.

3624 Market Street, Suite 508
Philadelphia, PA 19104
(215) 387 6784 Fax (215) 387 6959

Feb 15 1994

FINAL REPORT

PHASE II ONR CONTRACT # N00014-91-c-0204

ASSEMBLY OF A NEURAL ANALOG COMPUTER

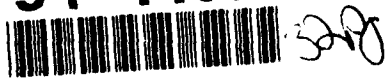


TABLE OF CONTENTS

1. Introduction

The aims of this project were the design and construction of a programmable analog neural computer. Original design work for the machine was initiated at the University of Pennsylvania with funding from the Office of Naval Research and the National Science Foundation. The design was patented and the patent was assigned to the University by the Inventors, Drs. Paul Mueller and Jan Van der Spiegel. Corticon Inc. obtained a license for the development and marketing of the technology and received SBIR Phase I and II awards for the actual construction of the machine. During Phase II additional funds were provided by the Benjamin Franklin Foundation of the State of Pennsylvania which provided matching funds for a sponsored research agreement between Corticon and the University of Pennsylvania under which Corticon gave $120,000.00 to the University for participation in this project.
Under the Phase I award a small protoype of the computer was built, demonstrating the feasibility of the approach. Details of the Phase I efforts were described in the Phase I report. This report describes the construction of the Phase II machine and its specifications. Because of delays in the manufacture of integrated circuits which were beyond the company's control, the completion date for the project received a 4 month no cost extension from DOD and the work was completed by Dec. 31 1993. On February 3rd, 94 the scientific Officer of ONR, Dr. Clifford Lau came to Philadelphia to inspect the machine and to see a demonstration of its operation. With minor exceptions detailed below the system performed in accordance with the specifications and Dr. Lau expressed satisfaction with the project. The local Press (Philadelphia Daily News) reported a brief account of this event.

The following sections (2-3) describe the Neural Computer, ancillary hardware, operating mode, operating software and projected applications. Section 4 deals with the development of Neural ASICS using the computer as a development tool. Corticon's efforts and plans for Phase III are summarized in section 5.

2. The Neural Computer

The development of the neural computer was motivated by the need for the unprecedented computational power required to solve real world problems in real time. Seemingly simple tasks such as seeing, hearing and motion, easily performed by any animal, have proved to be beyond reach for even the most advanced computers. Although many of the computational principles and algorithms used by biological brains are now understood, simulations of neural networks on digital machines have shown that such machines cannot provide the necessary speed. A sequential digital machine capable of matching the performance of the human brain would have to run at speeds approaching 10^{20} floating point operations per second (FLOPS) which is more than ten billion times faster than the fastest machines currently available. It is now understood that the astounding performance of biological brains arises from their ability to compute in a parallel and analog mode, both

of which transcend the iterative procedures of digital methods. Attempts to build machines which could match or surpass the brain's performance must take these facts into account.

2a Overview of the computer.

The neural computer developed by Corticon is modeled closely after the brain and is composed of electronic analogs of neurons, synapses and programmable interconnects. It can be expanded to any size without performance degradation. The current version contains over 800 custom VLSI chips and is capable of more than 10^{12} equivalent FLOPS.

The computer runs in analog mode but connection architectures, neuron parameters, synaptic gains and time-constants are set by a digital host computer that monitors the network performance and implements learning algorithms. The analog mode of operation enables truly simultaneous summation of many inputs at a single neuron, and the programmable synaptic time constants permit dynamic computations of temporal patterns as they occur in motion and speech.

The machine is intended for real-time real-world computations such as speech recognition, vision, robotics, control automation and other applications that require computational power and speed exceeding by many orders of magnitude the performance limit of digital machines. The system combines the unique inherent properties of neural networks, real-time or compressed-time speed for complex applications and features that are not available on neural network systems based on digital simulations.

The machine has several unique attributes: First, in addition to adjustable synaptic weights, it contains modifiable synaptic time constants, an indispensable feature for analog computation of time domain phenomena. Second, it is constructed from interchangeable modules with two dimensional symmetrical pinout, which allows a modifiable gross architecture in which the numerical ratios and positions between neuron, synapse and connection (switch) modules can be easily selected. Third, the network is expandable to any arbitrary size and can therefore be adapted to the complexity of the tasks. Finally, the direct coupling of analog and digital hardware adds the flexibility of digital methods to the machine.

2b Detailed description

For further details see references 1 - 12.

Architecture

The general architecture and a photograph of the Machine are shown in figs. 1 and 2.

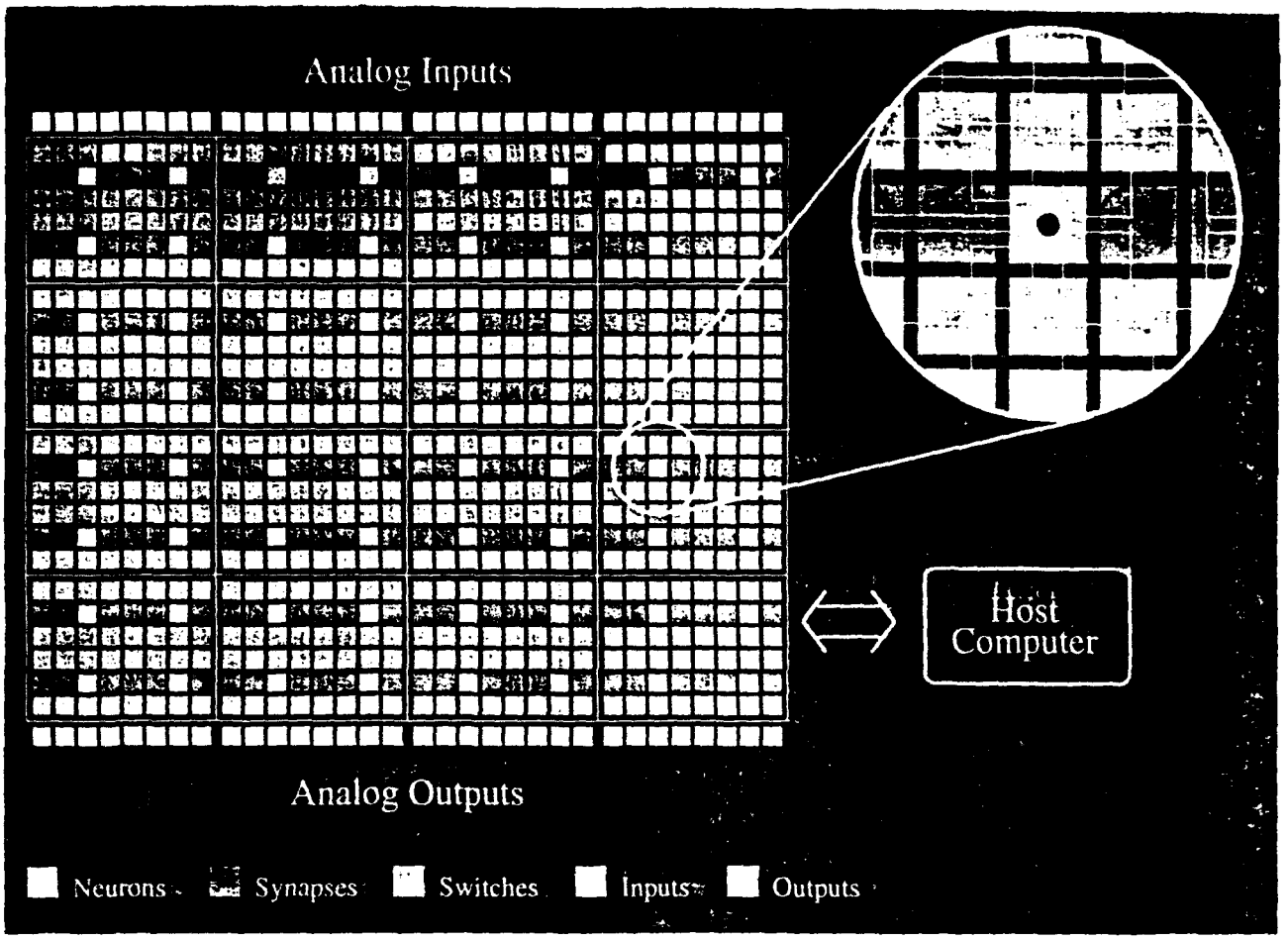


Fig.1 Block diagram of the Neural Computer

The machine consists of interconnected modules each of which contain on a VLSI chip an array of components (neurons, synapses or switches) and their control circuits. The current machine contains 64 neuron modules for a total of 1024 neurons each having 97 synapses. The symmetry of the connections between modules allows adjustment of the ratio between different modules and unlimited addition of modules. The insert shows the direction of data flow through the modules. Outputs from each neuron leave north and south and are routed through the switch modules east and west and into the synapse modules from north and south. There are also switch controlled lines on the neuron and synapse chips that bypass the synapses and neurons from north and south and connect to the switch modules. Input to the neurons through the synapses is from east and west. Power, digital control lines and multiplexed neuron outputs run east and west. The multiplexed outputs to the host computer are not shown.

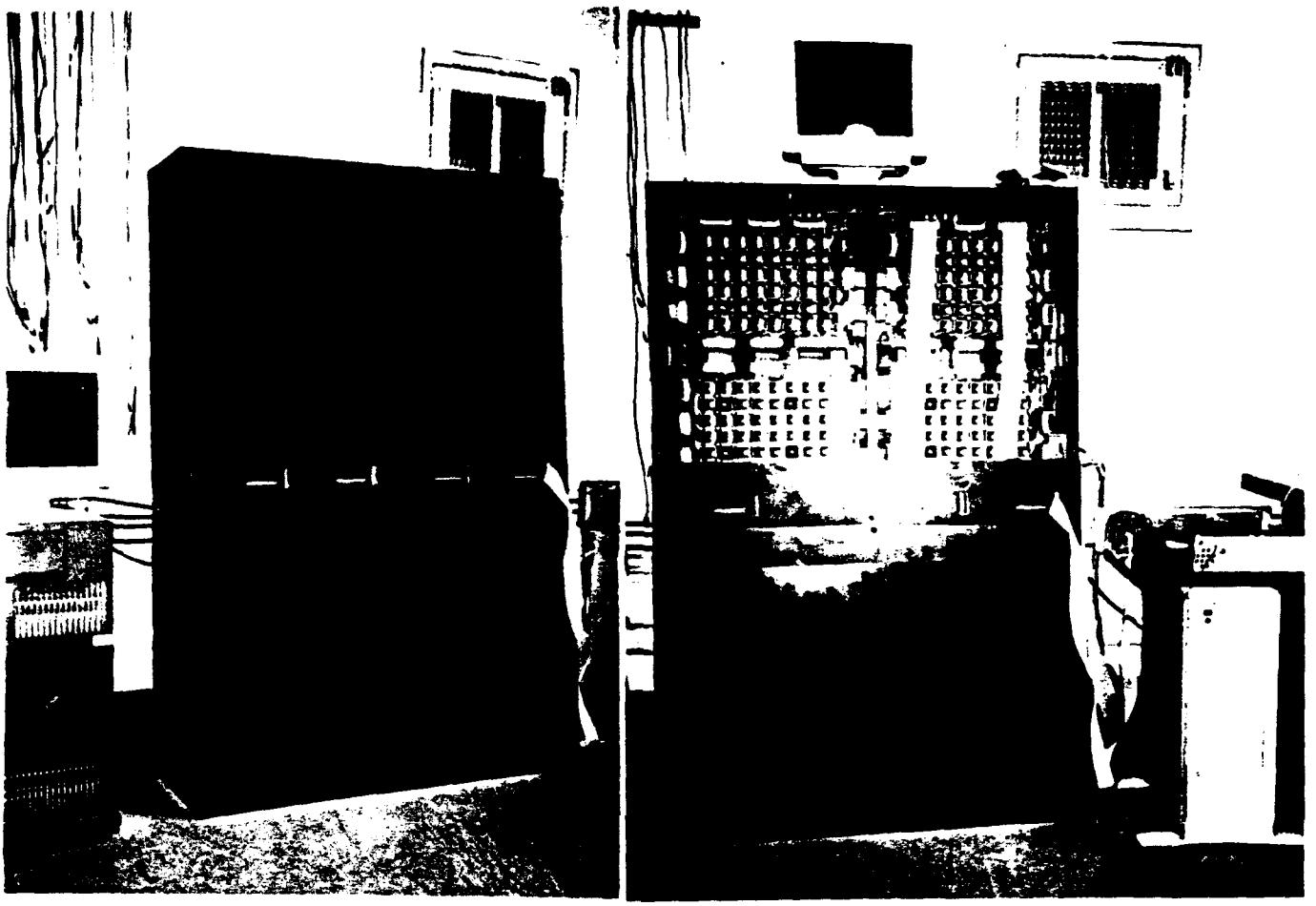


Fig. 2 Photographs of the Neural Computer. The digital host is shown at right.

In the current configuration, each neuron module contains on a VLSI chip 16 neurons, their control circuits and an analog multiplexer. The neurons are fed inputs from synapse chips, each of which consists of an array of 16 x 32 individual synapses. Each row of synapses receives inputs from up to 32 incoming lines and feeds the scaled signals to a neuron. In the current configuration each neuron is fed by three adjacent synapse chips for a maximum fan-in of 99 signals. The synaptic gains and time constants of each synapse are programmable over a large dynamic range. Provisions for additional inputs through cascading of synapse modules have been included in the design. Each switch module contains 4 arrays of 16X16 binary analog switches permitting an incoming signal to be routed to any or all of 32 lines. At the border of each array each line can also be turned off by a cut off switch to prevent the signal from being transmitted to adjacent arrays or chips. Adjacent switch modules are directly interconnected. Since the network design is completely modular, the machine can be expanded to any size and configurations involving different combinations of the modules are possible thereby expanding the neuron fan-out capability.

The VLSI modules are mounted on 160 pin planar chip carriers which are surface mounted in groups of 48 on specially designed surface boards. The boards are directly interconnected by high density connectors. The entire machine contains 16 individual boards for a total of 768 modules. Since the boards have symmetrical input and output configurations they may be connected in arbitrary topologies, e.g as a sheet, a cylinder, a cube, a torus or in other arbitrary geometries. Special geometries may facilitate the interconnections of certain network architectures.

The system specifications are summarized in Table 1.

System control

The digital host serves to program the network parameters. They include neuron parameters such as threshold and minimum output at threshold, synapse gain and time constant and the connection architecture (switch settings) all of which are set under software control via a custom digital control interface. The output of any or all neurons is also sampled through the analog multiplexer and analyzed after A/D conversion by the host computer. Processing within the network however is entirely analog. The external control allows for the implementation of various learning algorithms in which the host monitors the network's output and makes appropriate adjustments of synaptic and neuron parameters until the desired performance is achieved. As a result the system is able to learn particular tasks.

For gray scale visualization of the neuron activity, the multiplexed output of all neurons is fed to a separate A/D converter in the digital host and displayed as small rectangles on a terminal screen.

The connection architecture of the network is programmed by the user, allowing each neuron to be connected to any other neuron, but not to all of them. The synaptic weighting and transfer time constant (time for the signal to peak) of each connection is also user-determined. Software tools bundled with the system consist of an interactive graphics architecture design program with autorouting capabilities, a network controller, and a selection of learning algorithms. The controller includes a Software Interface and an interface for the standard "C" programming language to accommodate custom, third party software.

The full potential of the machine is realized in execution mode, especially in situations involving neural computation of dynamic systems i.e. situations in which time is a variable. Acoustical pattern recognition and the computation of moving images are prominent examples. But even for rapid recognition of stationary patterns such as high speed character recognition the machine would prove superior to sequential machines. In general, systems that require solutions of many simultaneous equations would benefit.

TABLE 1 SYSTEM SPECIFICATIONS

ARCHITECTURE	
Number of Neuron chips	64
Number of Neurons per chip	16
Total number of Neurons	1,024
Number of Synapse chips	192
Number of Synapses per chip	528
Total number of Synapses	101,376
Number of switch chips	512
Number of switches per chip	1,120
Total number of Switches	573,440
Total number of chips	768
PERFORMANCE	
Programming rates	
Switch setting rate	$2x10^6\ s^{-1}$
Synapse setting rate	$3.2x10^5 s^{-1}$
Time constants setting rate	$5\ x10^5\ s^{-1}$
Neuron parameter setting rate	$3.2x10^5\ s^{-1}$
Running Performance	
Maximal Equivalent FLOPS	10^{12}

The Component Modules

Three different modules form the basic building blocks of the machine.
Each module contains on a single chip arrays of either neurons, synapses or routing switches. Programmable synaptic time constants are located on the switch module and on the neuron module. The modules are packaged in planar chip carriers and have symmetrical pinouts permitting direct interconnects and arbitrary arrangements of the modules. Digital control signals are loaded serially either into a selected module, a group of modules or the entire machine.

The Neuron Module

Each neuron chip contains 16 neurons, an analog multiplexer, the control logic and 16 programmable time constant circuits permitting low pass filtering of the neuron outputs.
A block diagram of the neuron module is shown in Fig. 3.

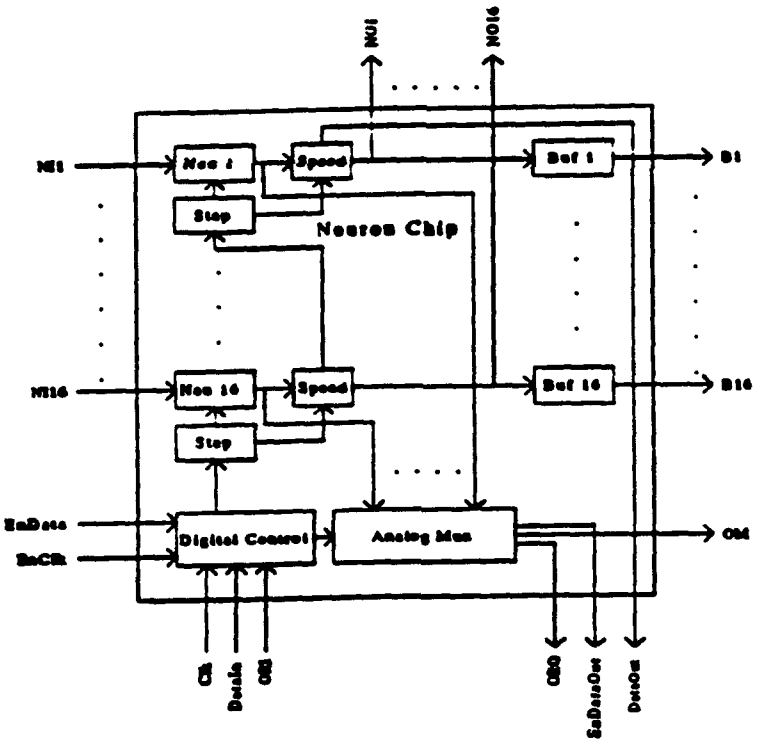


Fig.3. Block diagram of the Neuron module.

Inputs (NI) to each neuron come from synapse chips east and west outputs, (NO) go to switch chips north and south. The block labeled (speed) is a programmable time constant that controls output bandwidth and the blocks labeled (Buf) are independent buffered outputs that allow direct analog output to each neuron without going through the analog switch chips. The analog outputs are also multiplexed (Analog Mux) for A/D conversion and input to the host.

Input-output relations of the neurons are idealized versions of a typical biological neuron. The neuron transfer function are shown in Fig.4. Each unit has an adjustable threshold (bias), an adjustable minimum output value at threshold, Ex, a linear transfer function above threshold and a maximum output. Output time constants are selected either on the neuron chip or on the switch chips. The output has only one sign, positive or negative input polarity, i.e excitation or inhibition is selected at each synapse.

Neuron outputs can be routed via the switch modules in arbitrary fashion to different synapses throughout the net or to buffered output connectors at the front panel. There is also direct access to each buffered neuron output via a separate connector on each circuit board. In addition the neuron module contains an analog multiplexer that feeds the output serially onto a common line into a fast A/D converter located in the digital host. This permits the digital host to monitor the network performance and to implement learning algorithms by modifying synapse parameters and connection architecture on the basis of neuron outputs. The VLSI layout of the neuron chip is shown in Fig. 5.

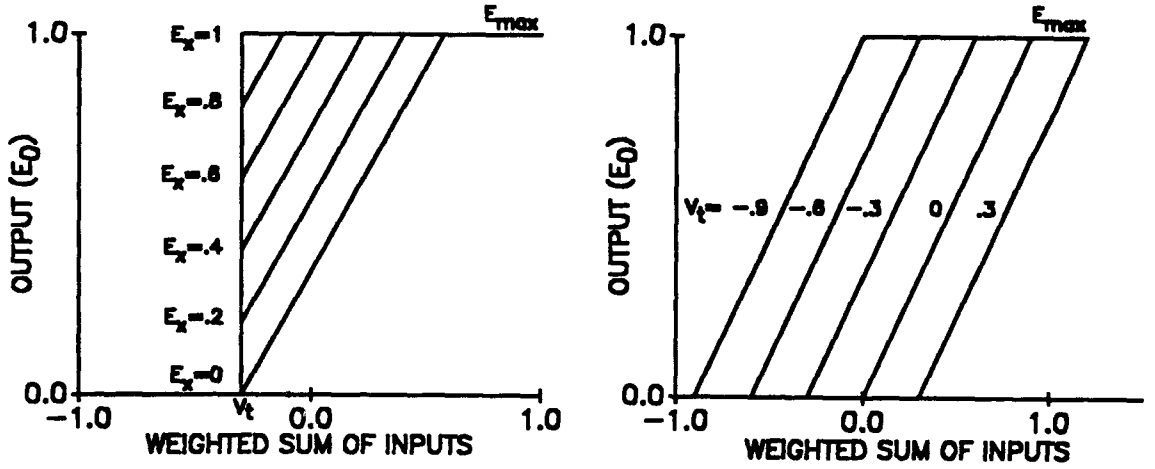


Fig. 4. Transfer function of the neuron.

Each unit has an adjustable threshold, Vt, a linear transfer region above threshold, an adjustable minimum output at threshold E_x and a maximum output, E_{max}. The adjustment of E_x and of V_t are shown in A and B.

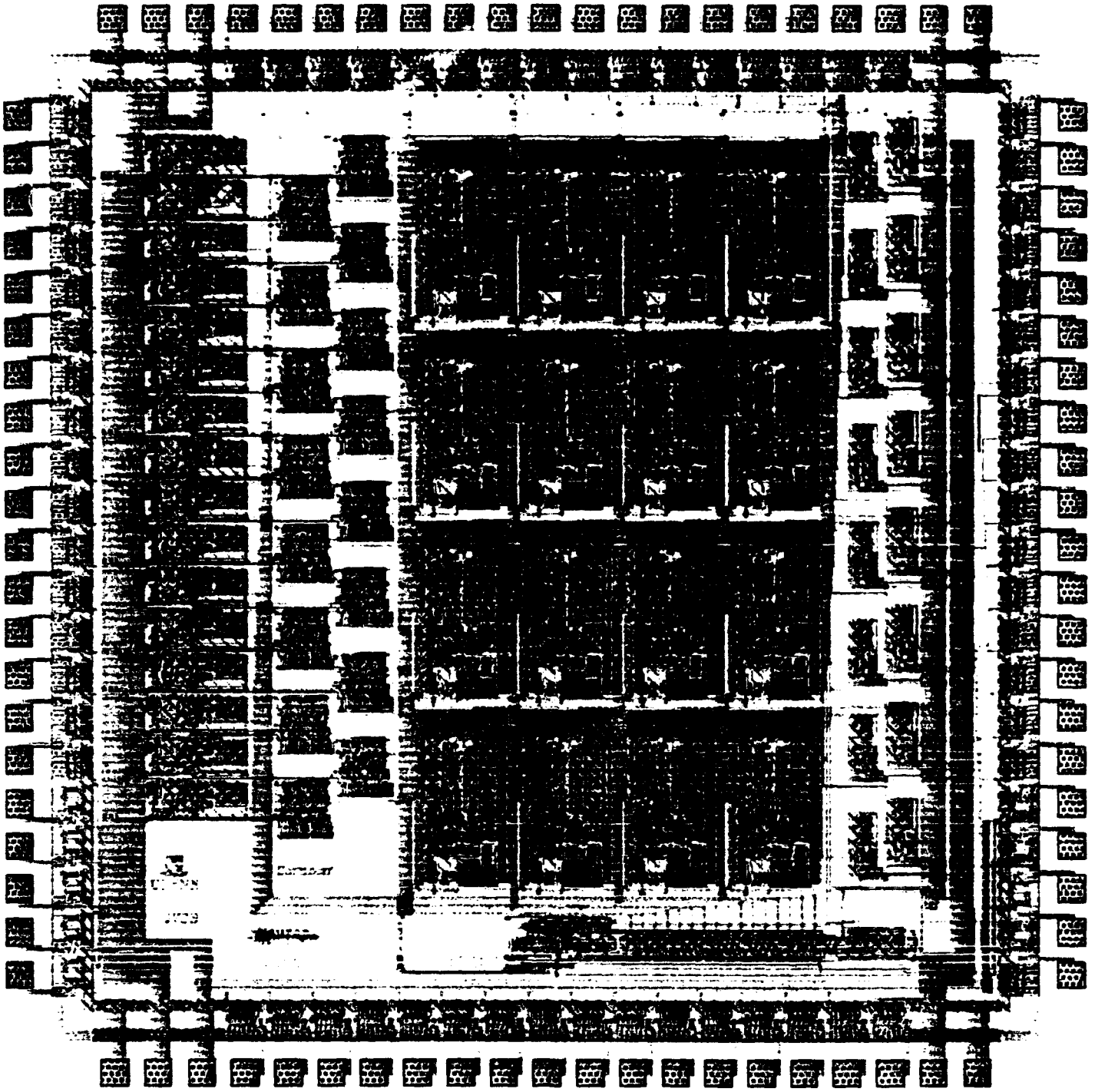


Fig. 5. VLSI Layout of the Neuron Chip.

The Neuron Properties

1. Threshold adjustment

The threshold of each neuron can be individually adjusted from the synapse chip via an extra synapse that is connected to a fixed input voltage. In this way the threshold can be biased in either direction. A neuron with a negative threshold bias produces output in the absence of inputs from other neurons. This feature allows negative logic operations and is also important for certain learning algorithms such as backpropagation.

2. The minimum output at threshold

Each neuron has an adjustable minimum output at threshold, E_X that can be set to any value between 0 and E_{max} by application of a current to pin E_X at the comparator circuit. This adjustment is the same for all neurons on one chip. It is an important feature which enables the neuron to perform logic, as well as arithmetic operations. In the limit the neuron can function either as a boolean switch when $E_X = E_{max}$, or as a summing amplifier when E_X is set to 0. Intermediate settings permit combined logic and arithmetic operations by the same unit. This feature is also found in biological neurons.

3. The input - output transfer function.

As currently implemented, the I/O transfer function is linear between E_X and E_{max}. This conforms roughly to the relation between average membrane depolarization and firing rate of a typical biological neuron tested in isolation.

4. Spiking neurons

Normally the neurons do not generate spikes as seen in biological neurons but transmit their output voltages as continuous variables. However each neuron can be programmed to generate spikes by low pass filtered feedback of inhibitory and excitatory connections utilizing the programmable time constants. In this way the entire machine can be transformed into an neural simulator useful for neurophysiological research.

The Analog Output Multiplexer.

The multiplexer provides the host computer with time segments of the outputs from all neurons. These data are used for monitoring the network performance and as a basis for the implementation of learning algorithms and adjustment of synaptic weights and or connection architectures. The multiplexer operation does not interfere with the actual operation of the net which is entirely in analog mode.

The multiplexer consists of 16 analog switches that connect the neuron outputs sequentially to a common output line. This output is buffered and provides a signal OM that is sent to an A/D converter over a common line. Multiplexer control is passed from one chip to the next allowing sequential reading of neuron outputs. The output signals are stored in the memory of the host computer. By using one or more fast A/D converters (1Mhz conversion frequency) the outputs of several hundred neurons can be read with msec time resolution which is fast enough to provide the host computer with time segments of the state of the network.

TABLE 2 NEURON SPECIFICATIONS

Process	1.5u CMOS
Operation	Current Summing
Outputs	0 to 3V
Transfer Function	Linear or sigmoid
Output Impedance	< 20 KOhm
Gain- Bandwidth Product	900 KHz
Settling time	1 us
Threshold Adjustment	Logarithmic, + or - , 7 bit resolution
Adjustment of E_X	0 to 3 V, 5 bits
Multiplexing rate	500 KHz

The Synapse Module.

Each synapse chip contains a 32 x 16 array of synapses. The weight of each synapse is set by serial input from the host computer via the digital interface and is stored at the synapse in a 8 bit local memory. The dynamic range of the synapse gains extends from 0 to 10 with 7 bit resolution, an eighth bit determines the sign. The gains are implemented by current mirrors that scale the neuron output after it has been converted from a voltage to a current.

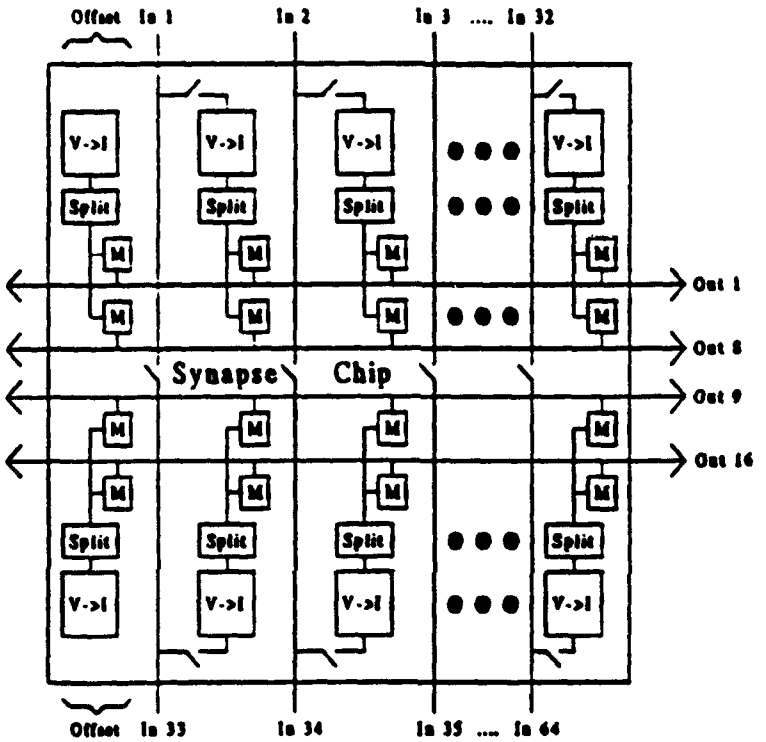


Fig. 6. Diagram of the synapse module. Each synapse gain is set by a 8 bit word stored in local memory (M). The memory is implemented as a quasi-dynamic shift register that reads the gain data during the programming phase. Voltage to current converters (V>I) transform the neuron output (In) into a current. (Split) are current mirrors that scale the currents with 8 bit resolution. The weighted currents are summed on a common line to the neuron input (Out). There are also switches that allow bypassing the synapses so that signals can be routed through the chip without going into the synapses.

The block diagram of the module is shown in Fig. 6. It consists of 2 arrays of 16 by 8 synapses and similar arrays of 8 bit memory elements which are mapped onto the synapse matrix. There are also 16 separate synapses per chip that serve to adjust the neuron bias. The chip has 32 input lines that are coming from neuron outputs, routed over switch modules. The inputs, which vary between 0 and 3 volts are transformed into a current by the V-I converters shown at the top and bottom of the diagram. Associated with the

converter is a current divider, which generates the required voltages to drive the synapse. Only one V-I converter and current divider is needed per column. There are 16 output lines, which carry the sum of the current outputs of the 32 synapses. These output lines are connected to the corresponding 16 inputs of the neighboring neurons or to the outputs of adjacent synapse modules. This arrangement allows to increase the fan-in by placing one or more synaptic chips adjacent to each other and connecting the outputs of one chip the outputs of its neighbor. In the current configuration of the machine there are 3 synapse modules per neuron module providing 96 inputs per neuron.

As tested the synaptic transfer function is linear from 0 to 3 V.

Although the resolution of an individual synapse is limited to 8 bits, several synapses driven by one neuron can be combined through switching, permitting greater resolution and dynamic range. Furthermore, mismatching of synaptic currents due to transistor differences can be compensated by this method. The layout of the synapse chip is shown in Fig. 7.

It might seem that the limited number of inputs per neuron restricts the computations that can be performed by any one neuron. However the results obtained by one neuron can be appropriately scaled and copied through a unity gain synapse to another neuron which receives the appropriate additional inputs. Furthermore, the symmetrical connections between modules allows for easy addition of synapse chips so that each neuron chips can served by more than three synapse chips, thereby increasing the number of synapses per neuron.

TABLE 3 SYNAPSE SPECIFICATIONS

Process	1.5u CMOS
Operation	Current Scaling
Weight control	Digital, 8 bits (logarithmic)
Dynamic range of weights (gain)	0 to + - 10
Resolution	floating point 3 + 4 bits[2]
Output Current Range	0 to 400 uA
Transfer Characteristic	Linear from 0 to 3 V
Download time	3us /synapse
Input Impedance	$> 10^{12}$ Ohm
Number of Synapses per Chip	528

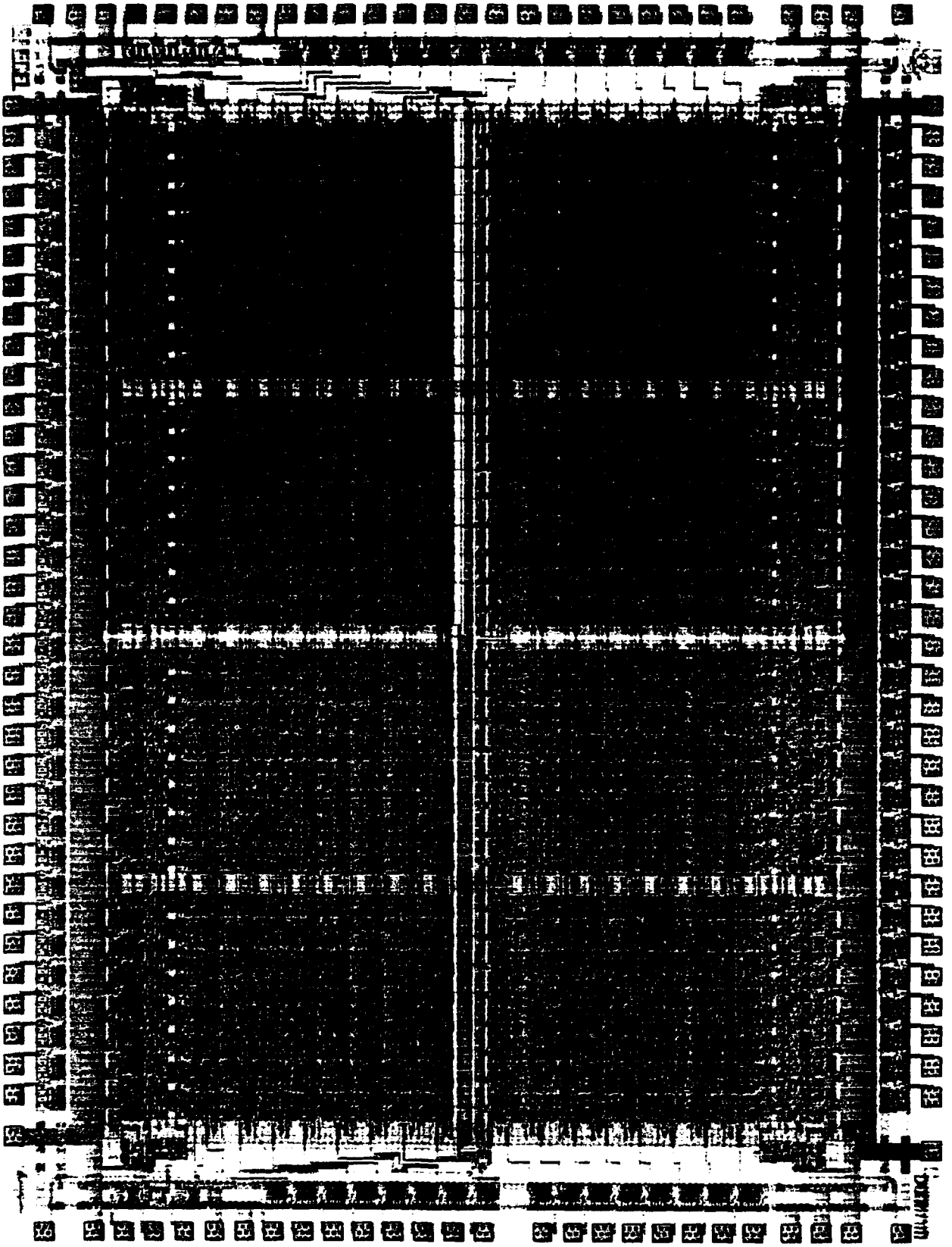


Fig. 7. VLSI Layout of the Synapse Chip.

The Switch Module

The switch modules serve to route the signals between neuron modules and the synapse modules thereby changing the connection architecture. Each module contains four 16 x 16 cross point arrays of analog switches which are set by serial digital input. Figs. 8-9 show block diagrams of the chip's major subsections. They consist of switching fabric, control logic and local memories that form a shift register. In addition the switch module contains 16 programmable time constants that are in series with the output lines.

Fig. 9 shows a block diagram of the switching fabric. Each square represents a one-bit switch control memory and analog switch. The control data is serially loaded into the shift register memory. The circles along the right and bottom edges also represent switches and memory. The switches that are in series with the horizontal or vertical signals allow the microprocessor to "cut" a horizontal or vertical trace in the switch chip. This allows the interconnection buses to be partitioned into several segments which increases the maximum number of obtainable connections. In addition to switches the modules contain circuits which control the time constants of the synapse transfer function.

The control scheme for the switch chip is identical to that for the synapse, neuron and time constants which makes it possible to program the entire computer serially through a single control line. The switch performance is summarized in Table 3. The layout of the chip is showns in Fig. 10.

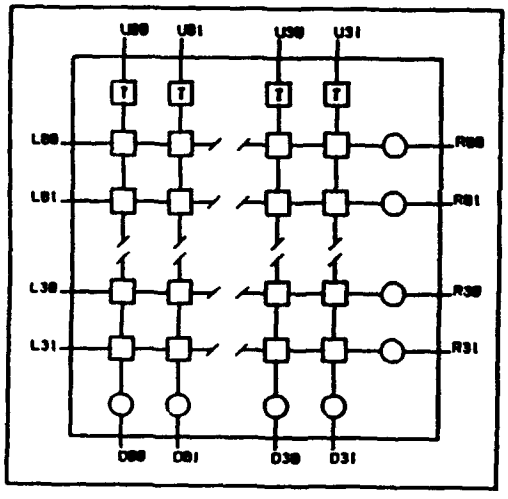


Fig. 8. Block diagram of the switch chip.

Squares and circles represent switch cells which connect the horizontal and vertical connectors or cut the conductors. The units labeled T represent programmable time constants.

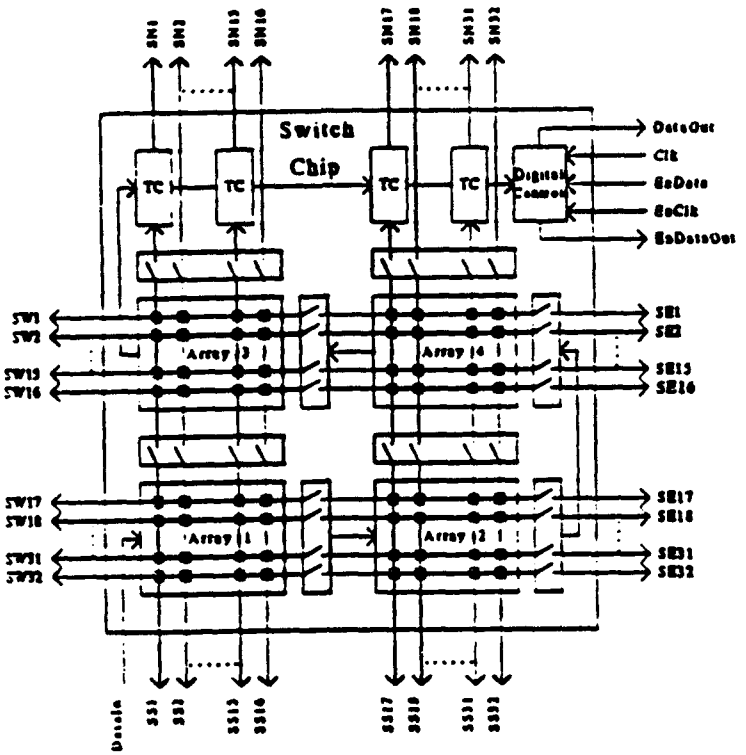


Fig. 9 Diagram of switching fabric.

The switches are used to route the analog neuron outputs. There are both cross point switches and cut switches. The cut switches segment the switching fabric into blocks. Switch settings are downloaded serially from the host computer into local memory. On every other line there are programmable time constants that are used to low pass filter or delay the signals at the synapse inputs for the computation of temporal patterns.

TABLE 4 SWITCH CHIP PERFORMANCE

Process	1.5u CMOS
Input capacitance	1pF
On resistance	< 3 KOhm
Download time	0.5 us/switch
Off resistance	> 1 TOhm

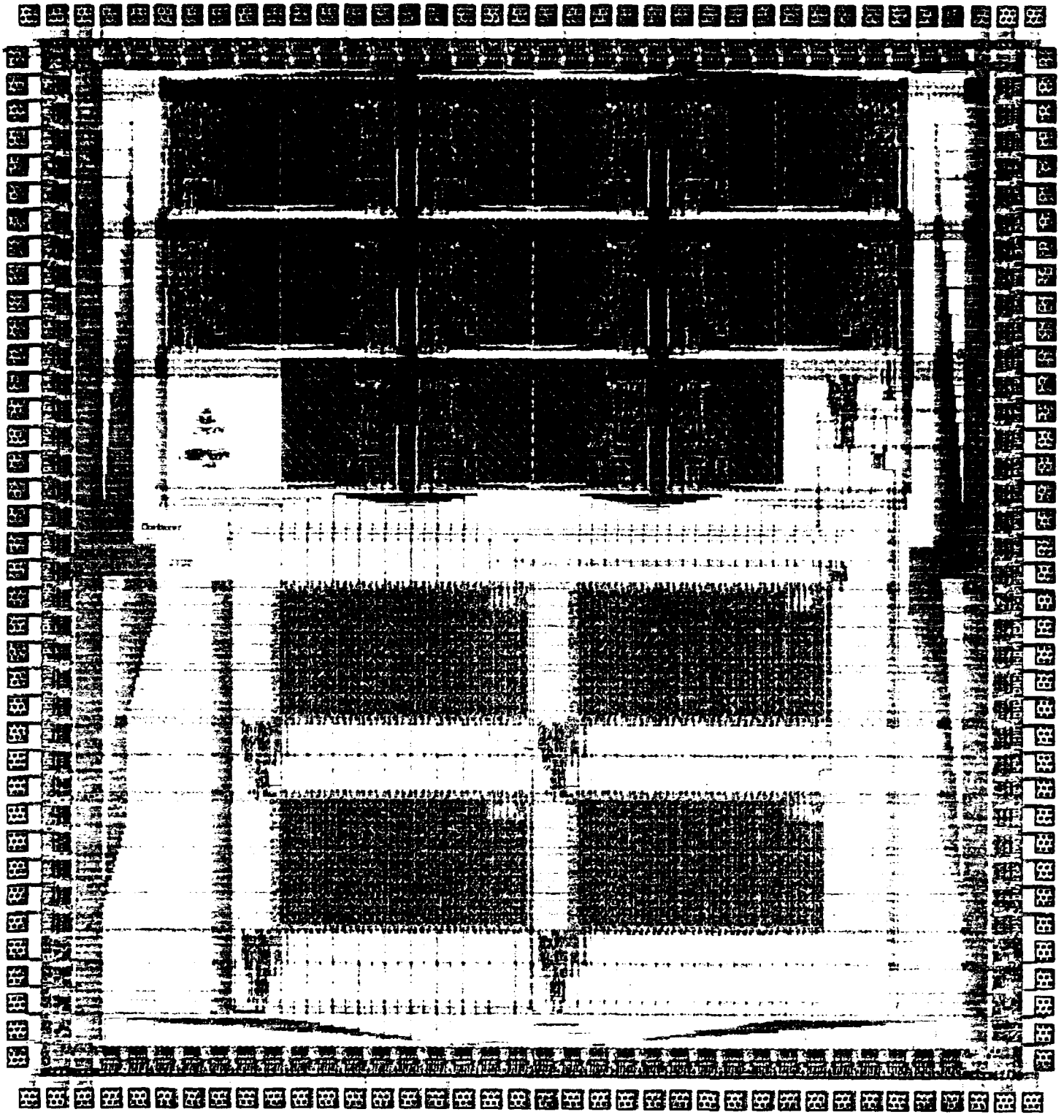


Fig. 10. Layout of the switch and time constant chip.

Adjustment of Synaptic Time Constants

For the analysis or generation of temporal patterns as they occur in motion or speech, adjustable time constants of synaptic transfer must be available. This is a very important aspect of neural computation, prominently found in biology but only beginning to be recognized. In order to deal with real world data, low pass filtering of the input signal to the synapse with 6 bit control of the time constant over a range of 50 us to 1s seems sufficient. By combining the low passed input with a direct input of opposite sign, both originating from the same neuron, time differentials of the signals can be generated giving the typical "ON" and "OFF" responses common in biological systems. In addition, serial connection of several time constants generates delay circuits for time delay computations.

The time constants are implemented by on-chip capacitors and programmable transconductance amplifiers. The time constant circuits are bi-directional and placed in series with the outputs of the switch chips at the points shown in Fig 8. Since not all synapse inputs need to have this feature, the circuits are placed on only one half of the number of output lines on the switch chip.

TABLE 5 TIME CONSTANT SPECIFICATION

Range	0.5ms - 0.5s
Digital control	5 bits log
Implementation	OTA, on-chip cap

Problems encountered during chip development.

As mentioned above, there was a long period of delays during the spring and summer of 1992 which were caused by a nation wide shortage of fabrication capacity at the VLSI foundries. As a result progress for the entire project was delayed by almost 6 months. In addition, the yields of the larger chips, especially the switch chips and their time constants was exceedingly low and varied from run to run. There seemed to be no design error, in fact several wafers had normal yields. As a result, we did not have enough switch chips with a full complement of functional time constants and had to make a compromise resulting in each board having only 3/4 th of the specified number of time constants. However, all boards are identical so that a particular network can be transposed into any portion of the machine. We have committed funds necessary to fabricate additional Chips and boards in order to bring the machine up to original specification. These chips are currently being packaged.

Control Hardware.

Fig.11 shows an overview of the different hardware components that load the network parameters and monitor the network performance. The network architecture is developed and stored on the Host computer. This data is then downloaded in serial form at 2-4 Mhz via a proprietary digital interface board (DI) located in the host. This board provides data buffers and clocks and permits both loading as well as verification of the loaded data. Another board (WB) is located in the neural computer and controls the distribution of the digital parameter data to selected sections of the network, individual boards or individual chips. It also controls the flow of the multiplexed analog neural outputs for A/D conversion by the high speed A/D board located in the Host.
There is a separate A/D board (LED) also located in the Host which converts the multiplexed neuron outputs at slower speed and provides continuous a gray scale representation of the neuron outputs as small blocks on a screen (DIS).

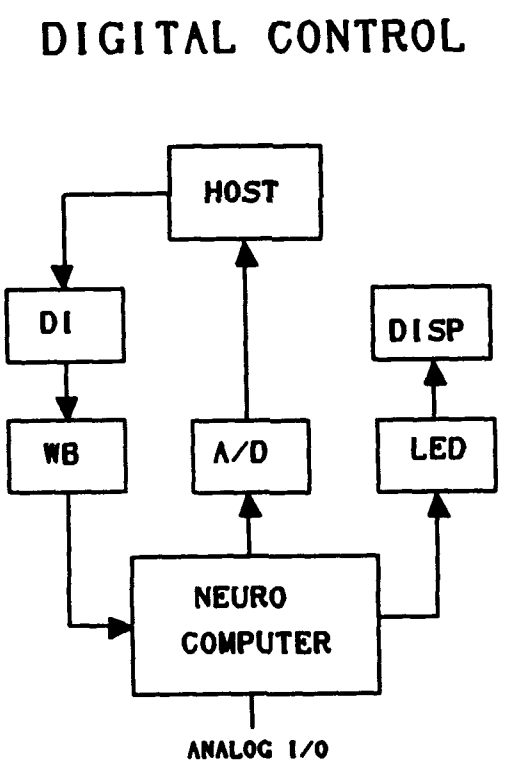


Fig. 11. Block diagram of the control hardware

Operating Software.

Fig. 12. shows a block diagram of the operating software. This software is stored in the Host and runs under UNIX. It generates the files of network parameters, loads these parameters into the neural computer and controls the acquisition of analog neuron outputs in digital form, their storage and use in the implementation of learning algorithms. It also controls the LED Monitor. There are two graphic programs for network design. The first allows the design of arbitrary conceptual nets as shown in Fig. 13, the second program presents a display of the physical network architecture, i.e. the neuron, synapse and switch chips and the available routing paths (fig. 14). The two programs are linked so that a conceptual net can be transformed into a physical net via the routing software and displayed as it is configured in the computer.

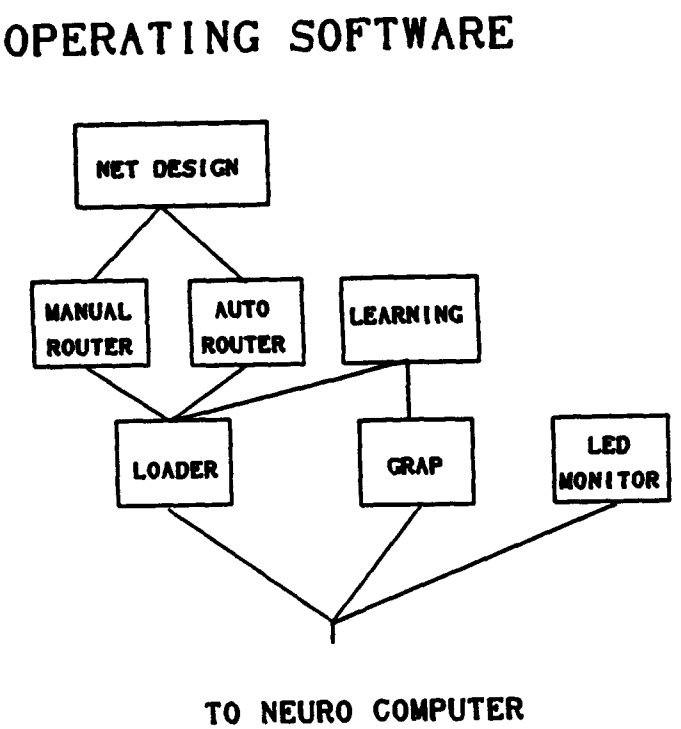


Fig. 12. Block diagram of the Operating software.

At the top level the desired network is designed by a graphics program that permits the design of arbitrary configurations and connections, entry of individual synapse weights, time constants and neuron parameters. This network configuration is then transformed by the autorouter into a file that contains the proper transform of the conceptual network into the physical network in the neural computer. This file is loaded via the loader into the neural computer through the DI and WB cards. At a lower level there is a manual router that generates the network architecture directly by setting the routing switches and synapse parameters under graphic control and displays an image of the actual network.

Separate programs (GRAP) control the operation of the fast A/D and the storage and display of the neuron activity. This data is used for the implementation of learning. The LED Monitor has separate software that controls conversion rates and display options.

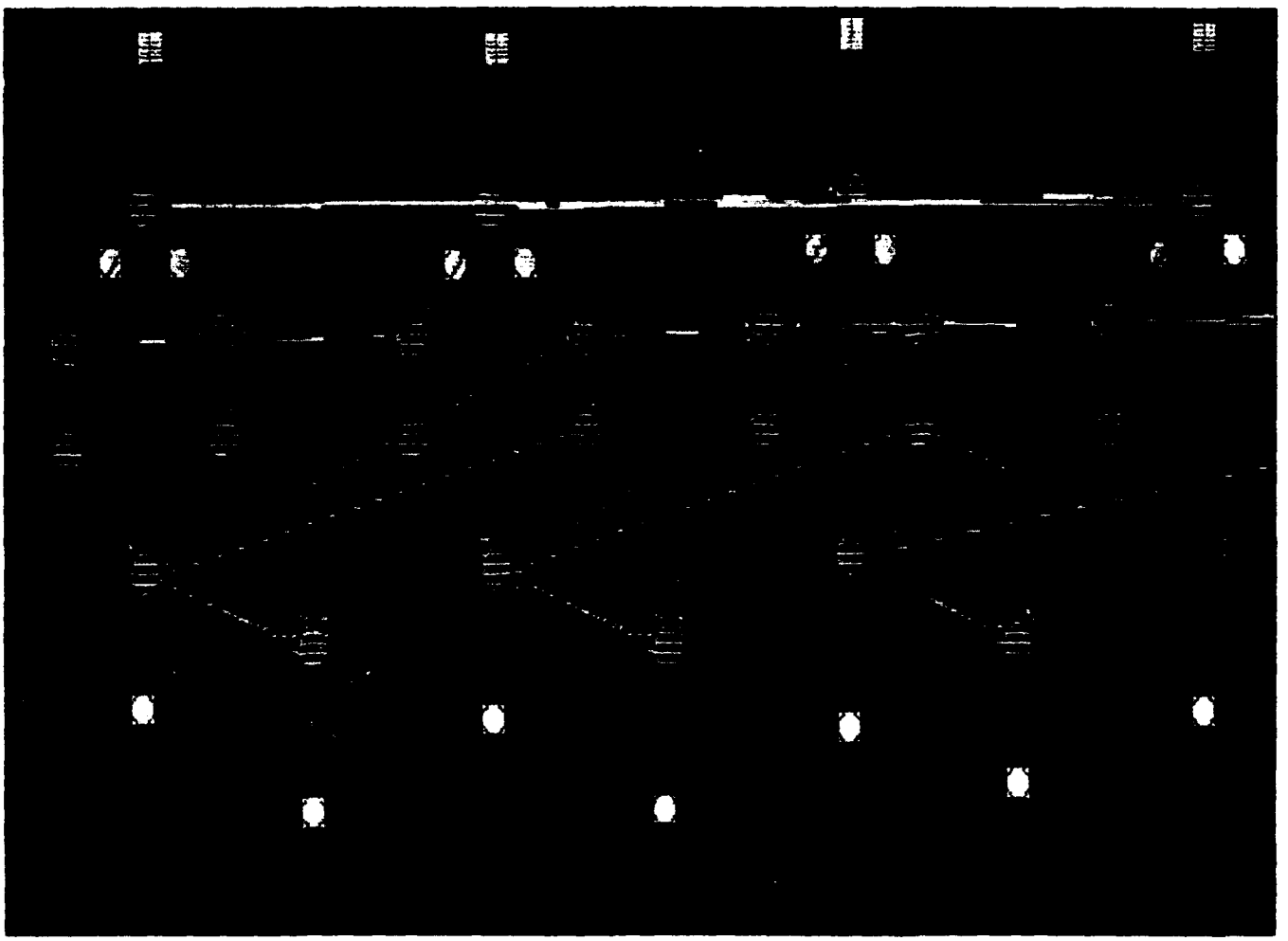


Fig. 13. Section of the conceptual design of a network for the primary decomposition of acoustical patterns. The primary neurons receive inputs from sixteen bandpass filters. These neurons are connected with mutually inhibitory connections in a center surround scheme with spatially decaying gains. They extract the maxima of the sound amplitudes at the different frequencies. The next stage extracts separately the temporal rise and decay of the sound amplitudes through the combination of delayed and undelayed excitatory and inhibitory inputs. The third and fourth stages compute the changes of frequency maxima and their direction through a combination of the second and third stage neurons. In essence they are motion detectors.

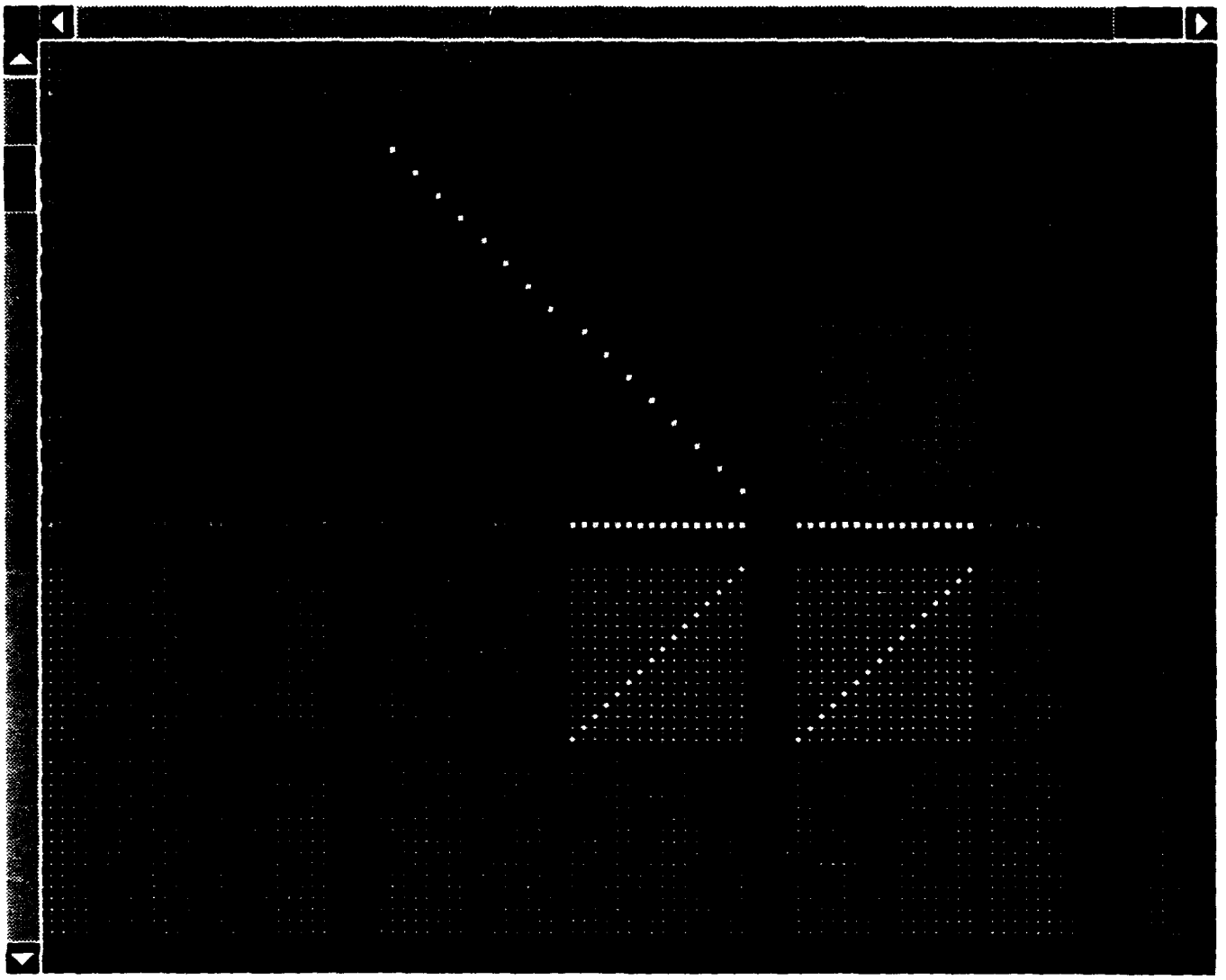


Fig. 14. Section of the graphic display of the physical editor and manual router showing a neuron chip, two neighboring synapse chips and three switch chips. Connections are made either manually by a mouse or from the conceptual design via the autorouter program.

3. Applications

Primary areas of application include real-world, real-time or compressed-time pattern analysis and recognition e.g. speech recognition, vision, target acquisition from fast moving vehicles, robotics, the design of dedicated neural circuits and the implementation of different learning algorithms.
The machine is also used to develop application - specific neural circuits (ASICs). Neural circuits developed on the machine can be directly transformed into ASIs from their neural components stored in VLSI libraries. Using the prototype as an emulation tool Corticon has developed analog neural circuits for real - time acoustical pattern decomposition applied to speech recognition and a chip for 2D real - time tracking of moving objects.
Corticon projects that analog neural networks and ASICs will play an increasing role as preprocessors, signal conditioners and classifiers in special purpose digital systems. The inherent fault tolerance, unprecedented speed, low cost and small size of neural analog preprocessing stages make them the ideal technology for future hybrid analog/digital systems.

Perhaps the most promising application of the machine lies in the area of speech recognition.

Corticon estimates that there is a potential market exceeding eventually $1 billion for computer systems that can recognize and transcribe continuous speech in real-time, are speaker independent, not limited by vocabulary and can be operated by speaking in a conversational voice, similar to how one would speak to another person. Such systems would find extensive use in any application that now relies on human operators for this task. However, the company believes that because of the enormous computational complexity, this problem will not been solved in the foreseeable future at reasonable cost by standard digital technology but that neural computers of the type the company has developed are sufficiently powerful to achieve this goal. A patent covering neural algorithms for acoustical pattern and speech recognition that have been implemented on the prototype is pending.

Development of algorithms for pattern identification.

As part of the Phase II effort, we have begun to investigate neural algorithms that can be applied to problems in ATR. Specifically we have developed methods for the identification of arbitrary visual patterns including human faces, vehicles, particle tracks in high energy physics or acoustical patterns.
The aim is to generate algorithms that are efficient, require no iterative computations, permit one-pass training and can be implemented by parallel neural hardware.
Our initial efforts have been quite successful. In general, the method involves several sequential steps.

The first is the computation of contrast by ON and OFF Center convolutions. This operation removes the average amplitude of the input pattern - loudness in acoustical patterns or illumination level in images. The second stage is the decomposition of the pattern into primitives by local neural circuits. Examples are orientation detectors for image decomposition or units that compute rates of frequency modulation in acoustical signals. These operations generate multiple sparsely populated matrices of neural activity that are characteristic for each pattern. At the third stage these sparse activity patterns are then decoded by proprietary dual convolution algorithms involving the sparse decomposition patterns and their negatives.

The synaptic weights from each stage to the next are obtained from a simple Hebbian learning rule on the basis of the activities in the previous stage. There is no backpropagation and no iterative gradient descent learning.

This method has been found to be very efficient and leads to reliable identification of arbitrary patterns even when buried in noise.(See Figs. 13 and 14.)

For the computation of visual images, we have designed and fabricated a silicon retina which computes the image contrast and forms the input stage to the neural computer. The subsequent image recognition processing or target tracking is then performed by the neural computer.

For acoustical pattern and speech recognition we have constructed a set of programmable bandpass filters that function as the input stage and decompose the sound into individual frequencies and provide a rectified and filtered signal of the energy in each frequency band. Further pattern decomposition into temporal and spatial derivatives as well as frequency changes and duration is performed by the neural computer. The final decoding of the patterns into neurons that respond selectively to specific patterns such as isolated or connected phonemes or diphones is performed by negative convolution techniques or by other "learned" nets.

Neural architectures that can implement these operations require many inputs per neuron and resemble the parallel fiber systems found in Cerebellum and Hippocampus, but the resemblance may be incidental. We are currently investigating the commercial and scientific potential of this method in areas of security (face identification), high speed optical pattern recognition (target identification, tracking, OCR, parts inspection) and speech recognition.

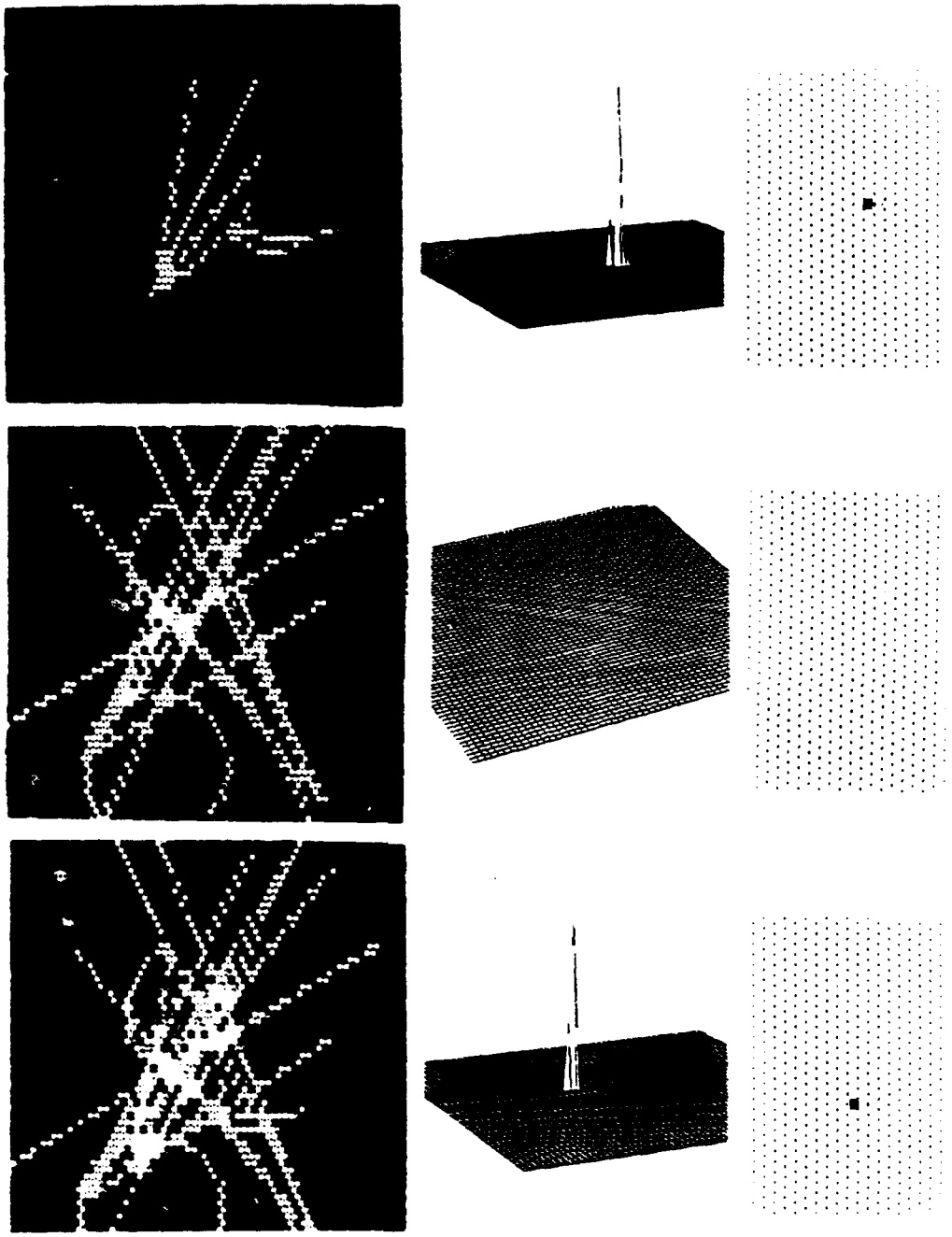


Fig. 15. Recognition of simulated particle tracks. The neural network was tuned by a special algorithm to the track pattern in the top left figure. The outputs from the tuned neuron array are shown in 3D representation to center and gray scale representation top right. The middle pattern is different and elicits no response At bottom the top and middle tracks are superimposed and generate a response. This decoding algorithm requires only one shot training and has been applied to more complex patterns such as human faces.

4. Development of Neural ASICs.

The machine is also used as a design tool to develop neural application specific circuits (ASICs) that provide the power and speed required for real-time applications.

As an example we have begun the development of a neural analog VLSI system that performs in real or compressed time the decomposition of acoustical patterns into selected stationary and dynamic pattern primitives and is able to decode the set of primitives generated by specific acoustical signals such as speech. The system incorporates to some extent the processing strategies of the biological auditory system and is intended to function as front end for the recognition of acoustical patterns including speech. It consists of directly interconnected ASIC modules, containing arrays of electronic analogs of neurons, synapses and synaptic time constants that perform a set of dedicated operations on the signals.

The circuits of the VLSI modules are designed and tested with the help of the neural computer used as an ASIC development tool. After a design is finalized, the layout of the ASIC chips is automatically assembled from the netlist and component libraries of the neural computer. These VlSI modules are unique because they contain programmable synaptic time constants and permit the real time computations of multiple coupled time differentials that are part of the dynamic patterns.

Some of the parameters to be extracted from the acoustical signal by the different neurons are the signal amplitude at a particular frequency, the rate of increase or decrease of the signal amplitude at this frequency, the rate of increase or decrease of the signal frequency, the amplitude modulation frequency, the signal duration and the signal sequence. Each neuron is tuned selectively to a limited range of these parameters.

Each module contains a single chip and performs these operations for a number of frequency bands over a limited frequency range. Since the component parameters such as synapse weights and time constants are programmable, all modules are structurally identical but their operating range can be adjusted from a digital processor or an EPROM. The number of modules in the system and the covered frequency range are determined by the desired frequency resolution and the specific application. For speech applications the range would extend from 100 Hz to 8Khz. Sonar or ultrasound applications would have different ranges.

The inputs and outputs are in analog form which permits direct interfacing with the neural computer for further data processing such as real time decoding and recognition of specific activity patterns. In addition the outputs are also available in high speed multiplexed analog form for A/D conversion and interfacing with a workstation or digital signal processor.

After the system has been assembled and tested it can be interfaced with the neural computer or a high performance workstation acting as a high speed front end for preprocessing.

5. Summary of Phase III activities.

As a result of the Phase II activities Corticon received a grant of $120,000 from the Ben Franklin Foundation of the State of Pennsylvania for further development of the Neural computer and related technologies. A portion of these funds were used to develop Neural VLSI chips for real time computation of visual motion and target tracking. The circuits were based on the VLSI components of the Neural Computer.

The Neural Computer received a patent issued to the University of Pennsylvania and the Corticon signed an exclusive License Agreement with the University of Pennsylvania on December 31st. 1992. The agreement covers the hardware technology of the Neural Computer as well as the software operating system. We consider the terms to be favorable and are pleased with the outcome of the negotiations.

The work has been presented at a number of national and international meetings. These included a DOD workshop on neural network applications in Alabama in September 1991, a Neural Network conference in Munich in October 1991, and a workshop on Hardware Implementation of Neural Networks held in Calgary from March 1-5, 1992 a Navy SBIR conference in April of 1992, an invited talk at the SPIE meeting in Orlando, April 1992 and a presentation at a National SBIR conference in Orlando in 1993.

The development of our Neural Computer was discussed in the Nov. 30th issue of the New York Times and in a brief section in "Electronic Design". In response to the NY Times article Corticon received a number of inquiries from investment bankers regarding venture funding. We plan to pursue these avenues after the machine has been completed. Due to unexpected low yields of the VLSI chips there has been a delay in the completion of the machine and a no-cost extension of the contract until the end of December 1993 has been granted.

Corticon is currently in negotiations with TMF Technology Management and Funding, 707 State Rd. Princeton NJ. which has expressed a strong interest in obtaining strategic partners in industry for our technology. TMF would raise $ 1.5 million over 9 months in return for 20% of equity in Corticon. We would prefer such an arrangement over direct VC capital because it would add additional resources from the prospective partner(s). Corticon shall decide by year's end if it shall enter into such agreement with TMF.

The Neural computer has been completed and Corticon plans to market the machine after it has undergone extensive testing and evaluation. In addition Corticon plans to use the machine as a development tool for neural ASICS in the area of speech recognition vision and control.

General Motors has expressed an interest in using the machine for the development of algorithms for drive train control using the programmable time constants as important control variables. If successful, the neural nets developed on the machine would be transposed into ASICS using the VLSI component libraries of the computer.

Researchers at the University are planning to use this approach for the development of ASICS for acoustical pattern decomposition and recognition. The NSF has expressed an interest in such a project and a large grant application is being prepared by the University. Corticon would provide the technology and its expertise.

Corticon shall also market the VLSI chips that comprise the modules of the neural computer. There are three different chip types (Neurons, Synapses and Routing Switches with programmable time constants) which can be used independently for applications in neural computation and other analog applications. The development of these chips is complete and they are currently available from stock.

6. Publications.

1. Mueller, P., Van der Spiegel, J., Blackman, D., Chiu, T., Clare, T., Dao, J., Donham, C., Hsieh, T.-p. and Loinaz, M. Programmable Analog Neural Computer and Simulator. In: **Advances in Neural Information Processing Systems 1**, D.S. Touretzky, ed., page 712-719, Morgan Kaufmann Publishers, San Mateo, CA, 1989.

2. Mueller, P., Van der Spiegel, J., Blackman, D., Chiu, T., Clare, T., Dao J., Donham, C., Hsieh, T.-p. Loinaz, M. A General Purpose Analog Neural Computer. Proceedings of IEEE and ICNNS International Joint Conference on Neural Networks, 1989.

3. Mueller, P., Van der Spiegel, J., Blackman, D., Chiu, T., Clare, T., Dao, J., Donham, C., Hsieh, T.-p., Loinaz, M. Design and Fabrication of VLSI Components for a General Purpose Analog Neural Computer. Analog VLSI Implementation of Neural Systems, IEEE Carver Mead, Mohammed Ismail, Eds., Kluwer Academic Publishers, Boston, MA. **1989.**

4. Mueller,P., Computation of Pattern Primitives in a Neural Network for Acoustical Pattern Recognition. **Proceedings of the IJCNN 1990 Wash.** I 149-151 .

5. Mueller,P.,Van der Spiegel,J., Agami,V. Blackman,D. Chance,P. Donham,C., Etienne,R. Kim,J. Massa,M., Design and Performance of a Prototype Analog Neural Computer. **Proceedings of the International Joint Conference on Neural Networks, Seattle 1991.**

6. Mueller, P. Van der Spiegel,J., Blackman,D. Chance,P. Donham,C.,Etienne,R. Kim,J. A Prototype Neural Analog Computer. **Proceedings of the Second Annual DOD Conference on Neural Networks. Huntsville Alabama. 1991.**

7. Van Der Spiegel J., Mueller P., Blackman,D. Chance,P., Donham,C., Etienne,R., Kim,J., Kinget,P. A Multichip Analog Neural Network. **Proc. Intern. Symp.VLSI Tech. Syst. Applications, VLSITSA 1991 pp 64-68.**

8. Van Der Spiegel, J., Mueller, P., Blackman D., Chance,P. Donham,C., Etienne ,R., Kinget. An Analog Neural Compuer With Modular Architecture for Real-Time Dynamic Computations. **IEEE Transactions On Solid State Circuits V 27 82 -92 1992.**

9. Mueller P., Van Der Spiegel, J. Agami,V. Blackman,D. Chance,P. Donham,C., Etienne,R. Flinn,J., Kim,J. Massa,M. Samarasekera S., Design of a Neural Analog Computer. **Proceedings of the 2nd International Conference on Microelectronics for Neural Networks, 1991 Kryll and Method Verlag Munich, 347-357.**

10. Mueller P., Van Der Spiegel, J. Agami,V. Blackman,D. Chance,P. Donham,C., Etienne,R. Flinn,J., Kim,J. Massa,M. Samarasekera S., A Programmable Neural Analog Computer. Neurocomputing, volume 4, 1992.

11. Etienne,R. Van Der Spiegel, Donham,C., Fernando, S., Hathaway, R. Blackman,D. **Mueller P.** A General Purpose Analog Neural Computer and a Silicon Retina for ReaL Time Target Aquisition, Recognition and Tracking. Proceedings of CAMP ' 1993.

12. Mueller, P. Van der Spiegel,J., Blackman,D., Donham,C.,Etienne-Cummings, R. Real Time Decomposition of Acoustical Patterns with an Analog Neural Computer. Proceedings of the SPIE, Orlando, 1992

Paul Mueller
Chairman

Peter Chance
Principal Investigator